Chiacchiere sintetiche

di **L. Nickhel**

(traduzione di **P. Capobussi**)

È finita l'era dei perenni distratti. Ogni minima operazione sarà accompagnata da una soave voce femminile che ricorderà di spegnere la luce dell'ufficio deserto, di attendere il risciacquo prima di togliere i panni, di inserire il proprio codice identificativo o di estrarre il pollo perchè è stracotto. Di contralto, con maggiore decisione, verremo avvertiti dalla voce virile che la temperatura ha raggiunto i 20° centigradi, o che è ora di cambiare dischetto, o che la radice quadrata di 9 è 3, anche se non te lo scrivo sul video. Segretari invisibili ed efficientissimi? Assolutamente no: un semplice chip di sintesi vocale, pilotato dal microprocessore di turno che stavolta si è anche preso in carico l'incombenza di gestire il periferico parlante.
In questo campo stanno rapidamente affermandosi dei chip VLSI, risultati delle intense ricerche condotte in questi ultimi anni riguardo i meccanismi che fanno dell'uomo un animale dalla chiacchiera facile.
Con pochi circuiti integrati sono ormai implementabili sistemi con il dono della voce.

Come produrre parole?

Pur essendo d'accordo con voi che a volte sarebbe meglio stare zitti e non aumentare il volume, già ampio, dei discorsi inutili, vediamo di identificare quali approcci sono adottabili per immagazzinare le parole necessarie ad un discorso.
Un sistema potrebbe essere quello di registrare su nastro ogni parola per poi riprodurla tramite dei classici circuiti di lettura e amplificazione dei segnali del nastro stesso. Un

WORD/PHRASE/LETTER	HEX	DECIMAL
THIS IS DIGITALKER	00	0
ONE	01	1
TWO	02	2
THREE	03	3
FOUR	04	4
FIVE	05	5
SIX	06	6
SEVEN	07	7
EIGHT	08	8
NINE	09	9
TEN	0A	10
ELEVEN	0B	11
TWELVE	0C	12
THIRTEEN	0D	13
FOURTEEN	0E	14
FIFTEEN	0F	15
SIXTEEN	10	16
SEVENTEEN	11	17
EIGHTEEN	12	18
NINETEEN	13	19
TWENTY	14	20
THIRTY	15	21
FORTY	16	22
FIFTY	17	23
SIXTY	18	24
SEVENTY	19	25
EIGHTY	1A	26
NINETY	1B	27
HUNDRED	1C	28
THOUSAND	1D	29
MILLION	1E	30
ZERO	1F	31
A	20	32
B	21	33
C	22	34
D	23	35
E	24	36
F	25	37
G	26	38
H	27	39
I	28	40
J	29	41
K	3A	42
L	3B	43
M	3C	44
N	3D	45
O	3E	46
P	3F	47
Q	40	48
R	41	49
S	42	50
T	43	51
U	44	52
V	45	53
W	46	54
X	47	55
Y	48	56
Z	49	57
AGAIN	5A	58
AMPERE	5B	59
AND	5C	60
AT	5D	61
CANCEL	5E	62
CASE	5F	63
CENT	60	64
400 HERTZ TONE	61	65
80 HERTZ TONE	62	66
20MS SILENCE	63	67
40MS SILENCE	64	68
80MS SILENCE	65	69
160MS SILENCE	66	70
320MS SILENCE	67	71
CENTI	68	72
CHECK	69	73
COMMA	6A	74
CONTROL	6B	75
DANGER	6C	76
DEGREE	6D	77
DOLLAR	6E	78
DOWN	6F	79
EQUAL	70	80
ERROR	71	81
FEET	72	82
FLOW	73	83
FUEL	74	84
GALLON	75	85
GO	76	86
GRAM	77	87
GREAT	78	88
GREATER	79	89
HAVE	7A	90
HIGH	7B	91
HIGHER	7C	92
HOUR	7D	93
IN	7E	94
INCHES	7F	95
IS	80	96
IT	81	97
KILO	82	98
LEFT	83	99
LESS	84	100
LESSER	85	101
LIMIT	86	102
LOW	87	103
LOWER	88	104
MARK	89	105
METER	8A	106
MILE	8B	107
MILLI	8C	108
MINUS	8D	109
MINUTE	8E	110
NEAR	8F	111
NUMBER	9A	112
OF	9B	113
OFF	9C	114
ON	9D	115
OUT	9E	116
OVER	9F	117
PARENTHESIS	A0	118
PERCENT	A1	119
PLEASE	A2	120
PLUS	A3	121
POINT	A4	122
POUND	A5	123
PULSES	A6	124
RATE	A7	125
RE	A8	126
READY	A9	127
RIGHT	AA	128
SS(PLURAL)	AB	129
SECOND	AC	130
SET	AD	131
SPACE	AE	132
SPEED	AF	133
STAR	B0	134
START	B1	135
STOP	B2	136
THAN	B3	137
THE	B4	138
TIME	B5	139
TRY	B6	140
UP	B7	141
VOLT	B8	142
WEIGHT	B9	143

Tabella 1 - La ROM National contiene 143 tra lettere, numeri e parole.

Chiacchiere sintetiche

semplice registratore abilmente pilotato da un microcomputer potrebbe adeguarsi benissimo allo scopo. È evidente, comunque, che oltre a tutti i problemi meccanici e di ingombro che un tale sistema comporta, il reperimento di una parola avviene in tempi estremamente lunghi, specie se confrontati con i tempi di elaborazione tipici del microprocessore che presumibilmente pilota il tutto.Il vantaggio che si ottiene consiste certamente nell'innumerevole quantità di parole differenti che tale sistema è in grado di gestire, limitate solamente dalla capacità di immagazzinamento del nastro a disposizione.

Un secondo approccio al problema è quello che consiglia l'utilizzo di convertitori analogico/digitali i quali, prelevato il segnale vocale, lo campionano ad una frequenza di circa dieci KHz e lo immagazzinano in forma numerica in PROM, dalle quali può essere riconvertito con una opposita trasformazione da digitale ad analogico. Purtroppo anche con una frequenza di 10.000 bit al secondo la necessità di memoria è elevatissima ed è già irragionevole pensare di immagazzinare più di una decina di parole.

Naturalmente è sempre possibile utilizzare una memoria di massa più ampia come potrebbero essere dei dischetti, ma certamente non è un sistema economico e neppure conveniente se si ha la necessità di attivare la voce abbastanza spesso.

Comunque è con tale approccio che molti hobbisti si sono potuti avvicinare, fino ad ora, a circuiti riconoscitori di voce per personal o a schede didattiche che prevedono esperimenti evoluti di conversione A/D e D/A.

```
10 'SYNTHESIZE ATTENTION
20 OUT12,58
30 D=50
40 GOSUB 180
50 OUT12,10
60 D=110
70 GOSUB 180
80 OUT12,73
90 D=20
100 GOSUB 180
110 OUT12,58
120 D=10:GOSUB180
130 OUT12,111
140 D=25:GOSUB180
150 OUT12,71
160 D=500:GOSUB180
170 GOTO 20
180 FOR X=1 TO D
190 NEXT X
200 RETURN
```

1

```
10 'SYNTHESIZE COMPUTER
20 OUT12,74:          'COM FROM COMMA
30 D=30:GOSUB150
40 OUT12,124:         'P FROM PULSES
50 D=30:GOSUB150
60 OUT12,52:          'U
70 D=70:GOSUB150
80 OUT12,139:         'T FROM TIME
90 D=20:GOSUB150
100 OUT12,81:         'ER FROM ERROR
110 D=30:GOSUB150
120 OUT12,71:         'SILENCE
130 D=500:GOSUB150
140 GOTO 20
150 FOR X= 1 TO D
160 NEXT X
170 RETURN
```

2

```
10 'SYNTHESIZE HELLO
20 OUT12,90:          'H FROM HAVE
30 D=20:GOSUB110
40 OUT12,43:          'L
50 D=80:GOSUB110
60 OUT12,103:         'LOW
70 D=70:GOSUB110
80 OUT12,71:          'SILENCE
90 D=500:GOSUB110
100 GOTO 10
110 FOR X=1 TO D
120 NEXT X
130 RETURN
```

3

```
10 'SYNTHESIZE YES
20 OUT12,52:          'U
30 D=25:GOSUB90
40 OUT12,81:          'E FROM ERROR
50 D=25:GOSUB90
60 OUT12,129:         'SS
70 D=500:GOSUB90
80 GOTO 20
90 FOR X=1 TO D
100 NEXT X
110 RETURN
```

4

```
10 'SYNTHESIZE NO
20 OUT12,112:         'N FROM NUMBER
30 D=30:GOSUB90
40 OUT12,46:          'O
50 D=60:GOSUB90
60 OUT12,71:          'SILENCE
70 D=500:GOSUB90
80 GOTO 10
90 FOR X= 1 TO D
100 NEXT X
110 RETURN
```

5

1 *Listato 1 - Sintetizzazione "attenzione".*

2 *Listato 2 - Sintetizzazione "computer".*

3 *Listato 3 - Sintetizzazione "hello".*

4 *Listato 4 - Sintetizzazione "yes".*

5 *Listato 5 - Sintetizzazione "no".*

L'analisi del problema ha portato ad alcune scoperte interessanti. Pochi anni fa si è arrivati a isolare degli elementi semplici, le combinazioni dei quali permettono di produrre qualsiasi parola o frase si desideri. La soluzione al problema di combinare tali elementi atomici del parlato sarebbe ovviamente banale per qualunque microcomputer, se non fosse che sorgono delle complicazioni dovute al fatto che questi singoli fonemi, meglio detti allofoni, devono essere modificati in relazione ai precedenti e ai successivi. Non solo, ma è pure necessario introdurre nei messaggi così prodotti degli elementi prosodoci che evitino emissioni mono-tono.

A tutt'oggi esistono differenti tecniche per eseguire il tutto e ancora, purtroppo, il parlato in uscita al sistema non è confrontabile con nessuna buona pronuncia nostrana. È comunque indubbio che con meno di 100 $ è già possibile procurarsi dei chip che possono virtualmente sintetizzare qualunque parola, spendendo solo minime quantità di memoria.

La maggior parte dei produttori di semiconduttori si è già da tempo lanciata su questa strada, e i prodotti del mercato cominciano ad essere presenti in numero considerevole. Diamo uno sguardo ad alcuni di questi, per poterci orientare in una successiva applicazione che potrebbe vedere coinvolto anche il nostro personal.

Texas Instruments

La Texas Instruments ha adottato il metodo della "predizione lineare", denominato LPC, per la cui descrizione rimandiamo alla bibliografia che si è voluta aggiungere in fondo, non essendo scopo immediato quello di sviscerare i metodi di sintesi.

Nella pratica hardware possiamo scegliere tra quattro chip: il TMS 5100 e il TMS 5110, il TMS 5200 e il TMS 5210.

I primi hanno il bus di controllo a parallelismo 4, attraverso il quale possono ricevere comandi da un microprocessore a quattro bit quale può essere, per rimanere in famiglia, il Texas TMS 1000. L'abbinamento a tale microprocessore ne fa strumenti d' uso generale, soprattutto adatti ad applicazioni a basso costo, utili in prodotti "consumer". A tale scopo i chip sono già comprensivi di un amplificatorino da 36 mW.

I TMS 52XX richiedono un bus a parallelismo 8, non supportano alcun amplificatore, ma permettono una qualità migliore del parlato. Avendo una linea di *READY* è possibile che vengano visti dalla CPU come memorie lente, che possono inserire appropriati stati di *WAIT* al ciclo di scrittura del micro. Una qualunque CPU a 8 bit può, inoltre, controllarlo sia in interrupt che in polling, essendo testabile dall'esterno lo stato del disposi-

Chiacchiere sintetiche

tivo stesso.
Entrambi i chip 5100 e 5200 adottano una memoria ROM esterna da 128 Kbit, che trasferisce al chip di sintesi i codici richiesti per la composizione delle parole e delle frasi. Per il solo TMS 5200 esiste anche la possibilità di utilizzare un buffer interno, il quale può venire caricato dal microprocessore di controllo, che a sua volta, quindi, può utilizzare della memoria dell'utente e non il chip specifico TMS 6100 fornito dalla Texas. Nelle versioni 5210 e 5220 si possono collegare fino a 16 chip TMS 6100 per consentire fino a 30 minuti di parlato.
Tra i vocabolari disponibili ne esiste uno dimostrativo, contenente frasi e parole particolarmente...inutili, ma è disponibile anche una memoria programmata per la generazione di oltre 200 parole tra le più usuali.

General Instruments

Questa casa commercializza un prodotto interessante, denominato SP 0256 il quale comprende il sintetizzatore, il microprocessore e ROM per circa 16 parole, tutto inserito sullo stesso chip. Eventuali espansioni sono sempre possibili utilizzando ROM esterna (SP 16, SP 32, SP 128).
Un set di tre chip, denominato VSM 2032, permette la generazione di 37 tra parole e frasi ed è direttamente interfacciabile ad un bus di microcomputer.

Votrax

La Votrax è la compagnia che ha introdotto i primi sintetizzatori di fonemi.
Ora presenta un chip denominato SC 01, il quale è inseribile in un set di altri chip per la produzione di 255 tra parole e frasi, con la possibilità di raggiungere le 1.000 parole o frasi in aggiunta a diversi effetti sonori.
Tali prodotti sono stati utilmente impiegati nella realizzazione di apparecchi per dare una possibilità di comunicazione a persone handicappate nella parola, applicazioni chiamate "phonic mirror handy voice" e che, per fortuna, fanno ben sperare in una utilizzazione non solo godereccia e commerciale di tali prodotti della moderna miniaturizzazione.
Destinato al mercato degli appassionati di robotica è indubbiamente da citare anche l'inserimento di un chip Votrax nel "personal robot" Hero-1, un aggeggio recentissimo che è in grado di offrire una tazza di caffè all'operatore, augurandogli buon compleanno nell'esatta ricorrenza.

Altri chip disponibili

Il mercato offre ancora nuove possibilità con la presenza di circuiti integrati realizzati in casa AMI, Philips, ITT, Telesensory. Nell'ordine possiamo scegliere tra S 3610, MEA 8000, UAA 1003 e Speech 1000, ai quali vanno aggiunti i relativi circuiti di memoria o di prelevamento delle codifiche necessarie.

Particolarmente interessante anche per un uso hobbistico risulta essere il prodotto National.

Il Digitalker

Con tale denominazione si indica l'insieme dei tre chip della National Semiconductor venduti sotto le sigle MM 54104 per il sintetizzatore e MM 52164 SSR1 e MM5 2164-SSR2 per la coppia di memorie ROM contenenti il vocabolario. In tale configurazione è disponibile un kit chiamato DT 1050.
La ROM contiene 143 tra lettere, numeri e parole (come in tabella 1), tramite le quali, con un opportuno software di "troncamento dei tempi di pronuncia" è possibile ricavarne molte altre.

L'interfacciamento con un personal

La soluzione hardware è semplice. La piastra del sintetizza-

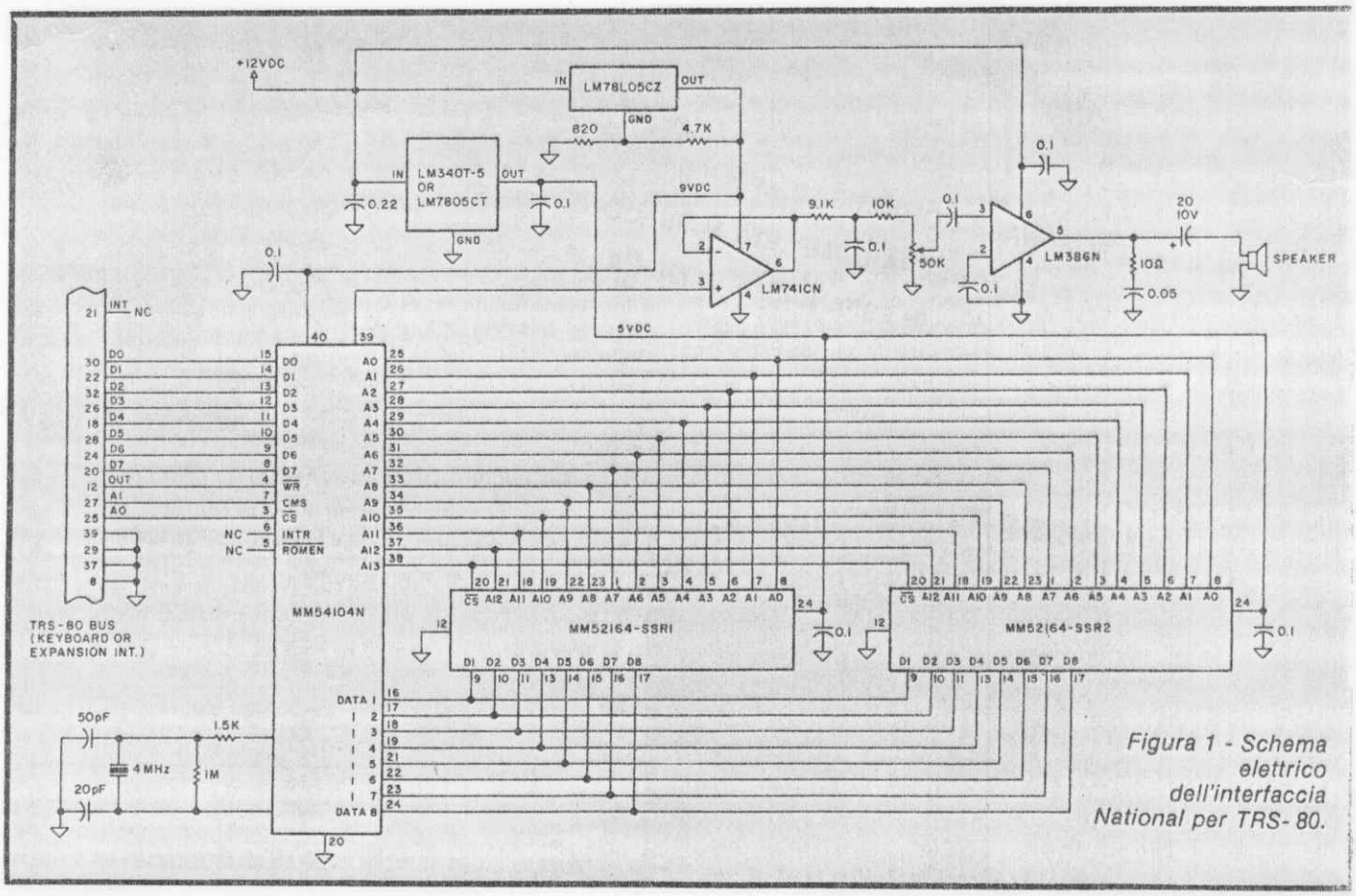

Figura 1 - Schema elettrico dell'interfaccia National per TRS-80.

tore può essere trattata esattamente come un dispositivo periferico di uscita, anche in mappa di memoria.

Le linee di controllo si riducono a un chip select *CS* e ad un piedino di write *WR*, entrambi attivi bassi.

Il sintetizzatore provvede al lancio di interrupt quando è pronto per una parola successiva, tramite l'uscita *INTR*.

La linea di command select *CMS* può essere tenuta bassa o, se volete resettare la linea di interrupt senza selezionare una nuova parola, va posta alta inviando un segnale di write.

Se si desidera controllare il tempo intercorrente tra una parola e l'altra si può tranquillamente evitare l'uso di *INTR*, mentre per risparmiare energia quando la ROM non viene utilizzata è sufficiente porre basso il piedino ROM enable (ROMEN).

Con un semplice amplificatore operazionale LM 741 è possibile realizzare un buffer a guadagno unitario che provvede ad inviare il segnale audio ad un amplificatore, nel nostro caso realizzato tramite un classico LM 386.

In figura 1 trovate lo schema di un circuito adatto allo scopo, realizzato per essere interfacciato ad un TRS-80, ma facilmente adattabile a qualunque altro microcomputer, essendo coinvolti solamente i segnali del bus dati, delle linee A0 e A1 del bus indirizzi e una linea di selezione *OUT*.

Il software necessario è molto semplice e non critico se ci si limita alla selezione delle parole contenute nel vocabolario standard. Con lo schema adottato una istruzione *OUT* alla locazione 0 permetterà la sintesi resettando *INTR*.

Una istruzione *OUT* alla locazione 2 permetterà invece di resettare *INTR* senza dar luogo all'inizio della parlata.

Quindi, ad esempio riferendosi alla tabella 1, se lanceremo dal TRS-80 una istruzione BASIC *OUT 0,76* otterremo in uscita la parola "danger".

La realizzazione pratica

Per quanto riguarda l'hardware solo poche raccomandazioni. Controllate accuratamente le alimentazioni prima di inserire il chip MM 54104, che essendo un NMOS è più delicato di altri chip. Per evitare premature distruzioni sarà posta cura nel maneggiarlo e nell'evitare di superare i 9 volt di alimentazione, essendo il chip realizzato per funzionare tra i 7 e gli 11 Volt.

Può anche essere una buona idea lasciare un certo spazio sul circuito stampato che si realizzerà, con qualche zoccolo libero, così che si possano eventualmente aggiungere delle ROM e dei semplici TTL che, deviando opportunamente i segnali di selezione, permettano di adottare vocabolari differenti.

Dalla tabella è facile ricavare una delle 144 parole disponibili, utilizzando opportunamente le istruzioni di *OUT* citate.

Volendo è anche possibile resettare il chip prima che la parola sia stata sintetizzata completamente. È così possibile agganciare diversi elementi di parole differenti per ottenere nuovi suoni e nuove frasi. Ad esempio se si blocca a metà la formazione della parola "comma" e si aggancia a questo tratto la parola "pound", è facilmente ottenibile la parola "compound" originariamente non presente sul vocabolario.

È evidente la criticità di un software di temporizzazione che blocchi l'emissione della parola al punto giusto, ma è relativamente semplice creare una routine che ci permetta di tentare con temporizzazioni differenti fino a trovare quella ottimale.

Il semplice studio delle routine proposte che permettono la sintesi di parole complesse, dal punto di vista vocale, quali sono "attention" e "computer", è immediatamente esplicativo del metodo utilizzato.

Con una spesa relativamente limitata è quindi possibile far parlare il nostro personal.

Con quale frase inizieremo le prove? Personalmente tenterei una favola: "C'era una volta Pinocchio e un chip di sintesi LPC, battezzato Grillo Parlante..."

Bibliografia

Elettronica Oggi n.1, Gennaio 1982; n.5, Maggio 1982.
Bit n.22 ,Novembre 1981
Microcomputing, Maggio 1981.

ALCUNI INDIRIZZI ITALIANI

National Semiconductor
Via Solferino,19
20121 Milano
Tel. (02)3452046

Texas Instruments
Nucleo Industriale
Cittaducale-Rieti

General Instruments
Via Anfossi, 32
20135 Milano
Tel. (02)5465514

AMI
Via Volturno,46
20124 Milano
Tel. (02)3452116

Philips
V.le Elvezia, 20
20052 Monza
Tel. (039)36351

ITT
Milano Fiori
20094 Assago (MI)
Tel. (02)824215

Arthur Luehrmann
Herbert Peckham

IL PASCAL PER L'APPLE

un manuale «attivo»
per imparare il Pascal in 30 ore

424 pagine, 18 500 lire

Zanichelli